Noname manuscript No.
(will be inserted by the editor)

Biological Knowledge-enabled BERT for Innovation in Biomimetic Design: A Case Study

Feng Sun, He Xu*, Yihan Meng

Received: date / Accepted: date

Abstract The biomimetic design provides an adequate solution to achieve an excellent design. However, the prototype space for biomimetic design is relatively large, and it becomes more and more challenging to find the required biological prototypes quickly and efficiently. In order to improve the design efficiency and enrich the means of biomimetic innovation, this paper proposes a biological knowledge-enabled bidirectional encoder representation from transformers (BERT) model to assist biomimetic design, namely BioDesign. We extract the biological strategies, functions and extract dimensional information from the Asknature as the data source. The linguistic expression model-BERT was used to recommend biomimetic strategies or functions combined with the biological strategies data. Finally, we take the biomimetic erosion wear resistant design of the valve core as an example and use the proposed BioDesign model to recommend biomimetic inspired functions. According to the recommended content of the BioDesign model, we obtained the erosion wear resistance strategies and designed the biomimetic structure. The erosion wear experiment proved the feasibility and effectiveness of the proposed method.

Keywords Innovation design · Natural Language Processing (NLP) · Biomimetic · Valve core

1 Introduction

Biomimetics is an interdisciplinary approach integrating biology and technology by transferring natures principles into a technological solution [1]. Meanwhile, the biomimetic design has been a hotspot of engineering research in recent years [2], [3], [4]. Exciting inspiration for designing new materials and structures can come from biological systems, where engineers can find ready-made solutions to solve problems [5], [6]. The focus of biomimetic design is to identify relevant biological strategies [7]. Generally, the common search process of identify relevant biological strategies for a given engineering problem can be concluded in three parts: consulting biologist, process natural language source of knowledge and using biological strategies in biomimetic database [8], as shown in Figure 1. However, biomimetic design as a biological knowledge driven design method is challenging because it requires biological and engineering knowledge [9], [10]. Although the unique design concept of biological evolution means that we can learn from natural structures and materials, biological solutions are often very different, which increases the difficulty of searching for target biological functions [11]. In addition, biomimetic design is also susceptible to subjectivity. To use the information of biological systems in new products, users need to understand these systems [9]. However, due to the gap of knowledge background, understanding biological information is a challenge for engineers [12]. Even in multidisciplinary cooperation, the understanding of biomimetic design in other disciplines is full of changes due to educational background and terminology differences. Therefore, finding matching biology or biological function for a given target through scientific and effective methods is crucial for biomimetic innovation.

The biomimetic design needs reorganizing and using biological information, so it is necessary to excavate the existing biomimetic design information fully. Science is a dynamic system of scientific progress [13]. Functional descriptions are gradually enriched with the deepening

Feng Sun, He Xu, Yihan Meng
College of Mechanical and Electrical Engineering, Harbin Engineering University, Harbin, China
E-mail: railway_dragon@sohu.com

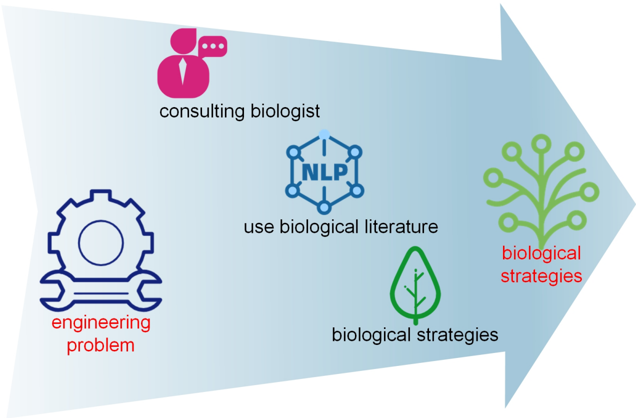

Fig. 1 The common search process of identify relevant biological strategies for a given engineering problem.

of biological and biomimetic research, and biomimetic language models composed of words, phrases, and matrices related to bionics are gradually expanding. Because of the similarities in the shape and function of biological expression, it is natural to associate natural language processing (NLP) with biomimetic design. Therefore, reading and understanding the biomimetic language model becomes more and more important. Although natural language has a well-defined vocabulary (approximately one million words in English) [14], there are still some differences in the vocabulary distribution from the bionic language model. How to better integrate the content of the bionic language model with NLP still needs extensive discussion. Recently, artificial intelligence has attracted much attention in language understanding [15], [16], [17], [18]. The artificial intelligence method can obtain and analyze and process massive data quickly and dig out the value contained in the data [19], [20]. By mining the value of data, the purpose of guiding the design process and improving work efficiency is realized. Natural language processing adopts the concept of rapidly improving pre-trained models with embedded transfer learning. The pre-training of the language model shows impressive results for natural language processing tasks [21]. BERT is a key technological innovation in the field of NLP. By pre-training, part of the NLP middle and downstream tasks are migrated to the upstream so that the model can achieve outstanding results in multiple natural language processing tasks. In particular, because BERT has achieved excellent performance in sentence NLP tasks [22], it is suitable for our purpose of understanding and searching bionic content. Pre-trained BERT means that it can be fine-tuned through an additional output layer and is suitable for the construction of the most advanced models for a wide range of tasks, such as question answering tasks and language inference, without requiring major architectural modifications for specific tasks [23], [24], [25].

The Asknature is the largest database of biologically inspired design [27], [28]. With the help of BERT and the characteristics of the Asknature database, we merge the biological strategies information in the Asknature with the BERT pre-training model. The biomimetic design method based on artificial intelligence assistance is proposed, and the key content recommendations of biomimetic design are given. The relationship model between the user's difficult-to-quantify needs and the design elements of the biomimetic model is constructed.

The essay has been outlined in the following way. Section 2 introduced the proposed method process. In section 3, a BioDesign model based on BERT is established with biological strategies and functions as feature input. According to the model results, the model output bionic function was discussed in relation to each other. In the section 4, take the erosion wear design of the valve core as an example, the effectiveness of BioDesign is verified through experiments. Section 5 provides conclusion and discussion, and presents our final remarks and suggestion of future work.

2 Proposed Method

In the problem-driven biologically inspired design, the starting point is the (engineering) problem for which suitable biological analogies are sought [29]. The important process of life design includes:

1. Searching for biological functions: Become more and more aware that all human beings, individuals and as a species, are part of nature, but independent, and respect the reciprocal relationship between all life through deepening connection.

2. Engineering mapping (simulation): The practice of scientific study and then copying natural forms, processes and ecosystems to create more regeneration designs.

Generally, the innovative design by biomimetic inspiration is a data driven innovative design method. The problem-driven (top-down) biomimetic design occupied the main part of the biologically inspired design [9]. The biomimetic design process of top-down mapping from top to bottom: questioning-biological search functional reference structural design. The biologically inspired design contains two important characteristic activities [30], [31]: searching for biological information and analogy transfer.

Carrying out problem-driven biologically inspired design requires that a design model is followed, as shown

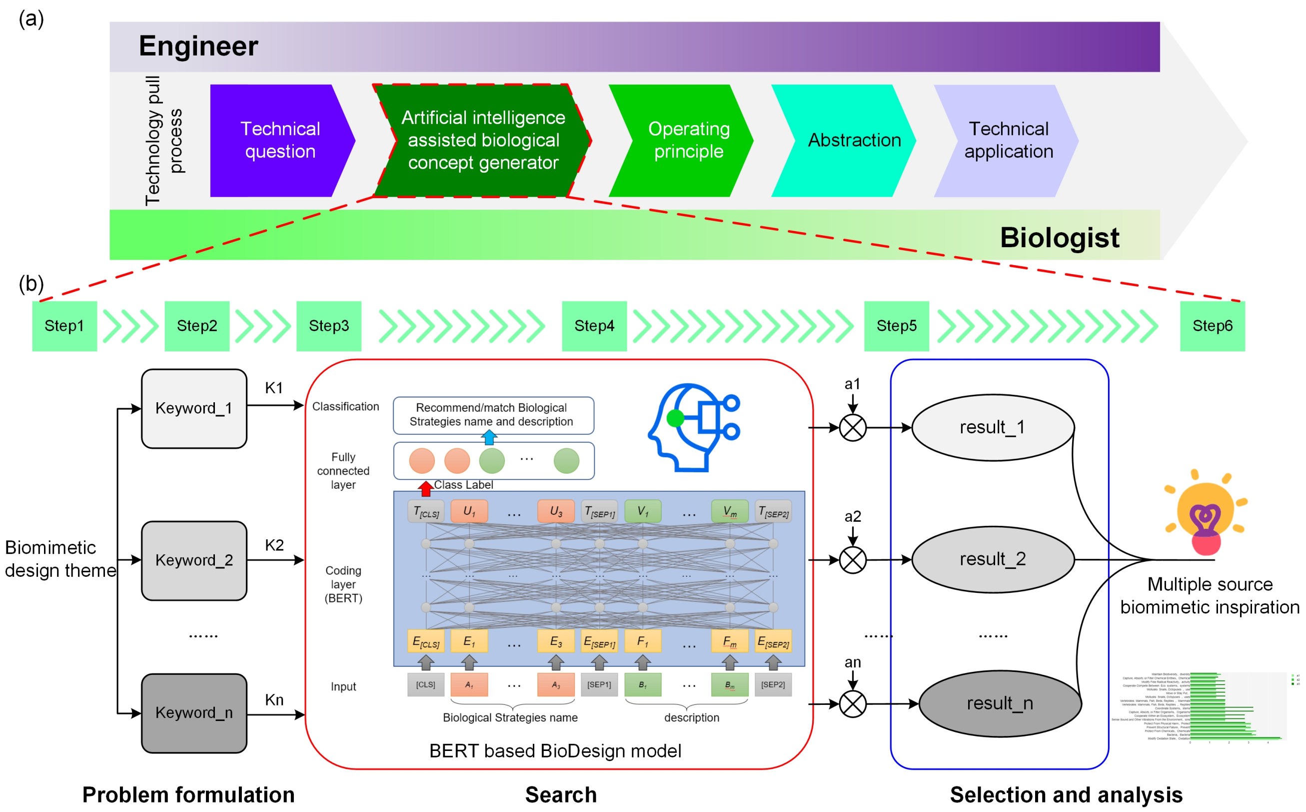

Fig. 2 (a)The top-down artificial intelligent aided approach (technology pull process) (adapted from [26]). (b)The proposed BioDesign biomimetic design framework. (Asknature: https://asknature.org, access in January 2021.)

in Figure 2a. Specifically, first, determine the requirements; then combine with part-of-speech analysis and other techniques to determine the bionic concepts or functions that meet the requirements. Then use the bionic concept or function as a design reference, extract feature information, and apply it to the actual design. Specifically, combining the steps of bionic design in the literature [29], [32], [33], the main steps of artificial intelligence-assisted bionic design are given. The process is as follows:

Step 1: Define goals and requirements

It is to set goals for implementing the bionic design to clearly understand its expected goals. Perform data analysis for those related biological functions or bionic designs that have clear answers.

Step 2: Data collection

In order to meet the goals and requirements identified in step 1, the appropriate data input that will be used for the analysis must be defined. This research aims to inspire bionic design, so it is necessary to obtain data related to the design function, such as biological description or functional description.

Step 3: Data understanding and preparation

The data cleaning process detects and corrects (or deletes) damaged and inaccurate data. This can be done manually or using data cleansing tools, which apply a set of predefined data validation and correction rules. Standardize and integrate data from different sources and different formats.

Step 4: Identify the feasibility of the problem

In this step, the designer conducts preliminary data analysis based on the extracted data. Then conduct preliminary analysis and verification of the survey results to determine the possibility of a solution.

Step 5: Analysis and modelling of bionic design

The solution obtained in the previous step is preliminary and requires in-depth and detailed analysis. The further analysis aims to discover the effectiveness of the bionic design. The model used for analysis can be data association, data classification, or simulation or experiment to evaluate the validity of the design.

Step 6: Apply biomimetic innovation design to bionic engineering

By combining bionic knowledge extracted from data analysis and engineering design, the bionic design of artificial intelligence-assisted analysis is realized. The proposed BioDesign model will be discussed in detail in the next section based on the above mentioned bionic design ideas.

3 BioDesign Model

3.1 BioDesign model dataset

The magical world composed of plants, animals and microorganisms has been researched and developed for billions of years. As young species, human beings are fortunate to get guidance from nature and inspire inspiration for innovative designs. The prerequisite for the accuracy of the combination of biology and engineering technology is collecting detailed information. Asknature (https://asknature.org/) summarizes examples of biomimetic innovation or engineering design inspiration from nature. Therefore, Asknature data is selected as the research data source, and the relevant content of biological strategies in Asknature is extracted through a self-developed python program. Obtain attribute parameters such as biological or bionic engineering cases and functional information, and the data processing format is shown in Figure 3(a-3). Furthermore, store the data in Excel format to be extracted during BioDesign model training.

3.2 BioDesign model based on BERT

Our BioDesign is initialized from BERT, which is an important recent development in universal language representation. BERT [34] uses bidirectional masking language modelling, in which a small part of the words are masked, and the model predicts the masked words. The use of BERT allows two-way training of the language model. Compared with the one-way language model, the two-way language model has a deeper understanding of the language context and language process [35]. In two-way language modelling, the model looks at all surrounding contexts that mask the tags. BERT refers to the learning method of the human brain and learns on large-scale text predictions, making the model have a strong text representation ability. Only simple fine-tuning in downstream tasks can achieve good results. BERT has learned large-scale corpus, which contains rich semantic knowledge, and it has certain advantages to transfer it to the field of biomimetic knowledge. This paper uses a BERT parameter model based on large-scale corpus pre-training and fine-tunes it on the biomimetic design and biological function data set. The data set provides the name, type, function and biomimetic design related to the creature. The BioDesign dataset contains three parts: biological strategies from Asknature, Google dictionary and English Wikipedia, as shown in Figure 3a.

3.2.1 data pre-processing

The BioDesign model uses the same data preprocessing as the BERT model. BERT takes a sequence of word tokens as input. [CLS] stands for the beginning of a sentence. For an input sequence composed of two sentences, the two sentences are separated by another special mark, marked as [SEP], representing the sentence's end. The input format of BioDesign is [CLS] Biological strategies name [SEP] corresponding description [SEP]. The input samples can be words, phrases or sentences. The length can be chosen arbitrarily but should not exceed the length of 512 samples. As shown in the input layer in Figure 3b.

3.2.2 model structure

The BERT-based BioDesign model structure is shown in Figure 3b. First, in the input layer, vectorize the biological strategies name and description text after data preprocessing to form an input format that the BERT model can accept and enter the BERT coding layer. This layer presents the BERT model structure in detail, where E and F are the input of this layer and the output result of the input layer. After entering the coding layer, the vector is subjected to bidirectional Transformer coding to obtain a comprehensive vector coding based on contextual semantics. The output value of the fully connected layer of the BERT model is used as the input of the next layer of the network to enter the matching layer. This layer uses the Softmax classifier to classify the results to obtain the matching degree between the texts. The BioDesign model of the BERT-based biological strategy classification task is shown in the Figure 3b. From the model's perspective, the figure's biological strategies name and description are independent in their respective fields. They have different data and labels, but the model space of the trained model is shared.

In the training process, the cross entropy is used as the loss function:

$$L_{similarity} = -\frac{1}{N}\sum_{i}^{N} y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i) \quad (1)$$

where, y_i is the true label, $\hat{y}_i$ is the predicted probability. In the prediction, the higher probability is selected as the resulting output. Based on the text matching model, input the category label [CLS], and train the similarity between the input text and the biological description through the fully connected layer.

Because the pre-training phase requires a lot of computing resources and time, this article uses Google's open source BERT pre-training model (bert-base-uncased)

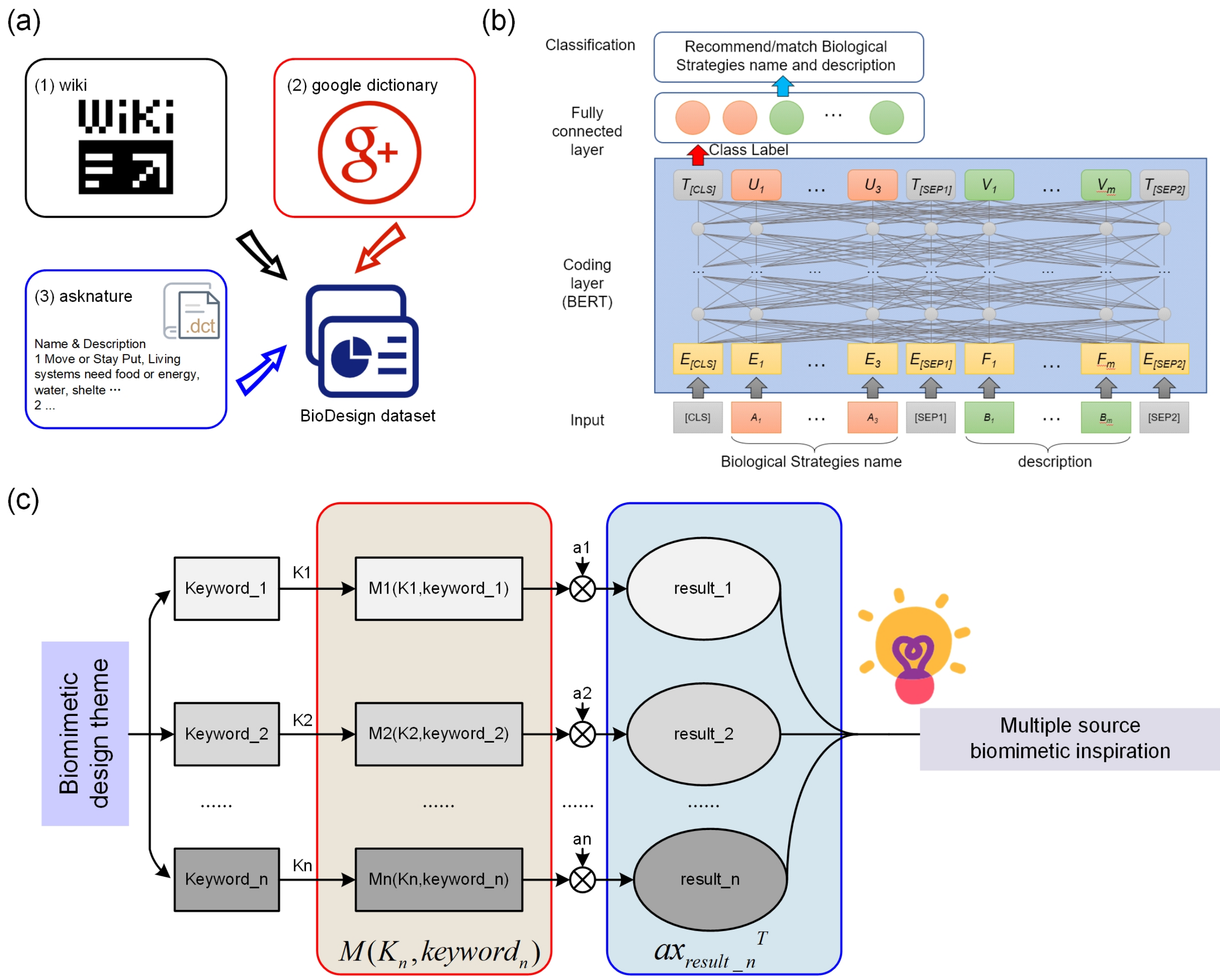

Fig. 3 BioDesign model :(a) BioDesign dataset; (b) Illustration of BERT for Biological Strategies classification tasks (adapted from [34]); (c) correlation calculation of multiple source biomimetic inspiration based on BioDesign model.

directly. The model weights can be directly loaded in the code for training. The fine-tuning stage is trained on the data set that matches the user's needs with the biological strategies name and description. The fine-tunes the parameters of the overall model composed of the pre-trained network and the fully connected layer. Through data training, the loss is calculated, and parameters are updated in the reverse direction. The purpose is to adjust the parameters of the model so that the model fits the data better.

Table 1 BioDesign model training parameter setting.

Hyperparameter	value
learning rate	5E-5
epoch	100
batch size	64
optimizer	Adam
activation function	tanh
test-size	0.1

3.2.3 Model training parameter

The BERT-based BioDesign model built in this article has the following steps:

(1) put the standardized biological strategy data after text preprocessing into the input layer.

(2) put the results of the input layer into the BERT pre-training model training.

(3) put the word vector obtained from the BERT coding layer into the fully connected layer and fine-tune to adapt the model to different learning tasks. The training model parameter table is shown in Table 1.

3.2.4 Model training metrics

According to the official evaluation system, for each task, the *precision*, the *recall* and the $F1$ score are computed [36]. In addition, accuracy is also used as an evaluation index in the process of model training. The metrics calculations above are derived from the following derivation.

(1)*precision*

$$precision = \frac{TP}{TP + FP} \tag{2}$$

where, TP(True Positive) indicates the number of correct results predicted, FP(False Positive) indicates the number of correct results predicted

(2)*recall*

$$recall = \frac{TP}{TP + TN} \tag{3}$$

where, *recall* denotes the number of correct results predicted by all models, including the original number of correct results and incorrect results. TN(True Negative) denotes the number of correct results completely found

(3)*accuracy*

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where, FN(False Negative) indicates the number of correct results predicted as wrong results

(4)$F1$

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{5}$$

$F1$ reconciles the contradiction between accuracy rate and recall rate, taking into account the two indicators' value.

The calculation results of the BioDesign model are shown in Table 2, where Label0 represents the mismatch between the name and description of biological strategies, and Label1 represents the match. The goal is to make the model learn biological strategies score of whether name and description match the training results of BioDesign model is shown in the following table.

Table 2 BioDesign model training results.

task	precision	recall	accuracy	F1
label0	0.72	0.67	0.71	0.69
label1	0.71	0.75	0.71	0.73

3.2.5 Word similarity

Word similarity is a value ranging from 0 to 1 according to the distance of the word in the thesaurus. The higher the similarity is, the closer the similarity is to 1. The similarity between a word and itself is 1. In this paper, vector cosine values are used to determine the similarity of two words. Define word similarity $sim(\mathbf{x}, \mathbf{y})$ as:

$$sim(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{n} x_i \times y_i}{\sqrt{\sum_{i=1}^{n} (x_i)^2} \times \sqrt{\sum_{i=1}^{n} (y_i)^2}} \tag{6}$$

where, the $sim(\mathbf{x}, \mathbf{y})$ are the compared word vector. $sim(\mathbf{x}, \mathbf{y})$ value range is $[0, 1]$. When two word vectors are synonyms or identical words, word similarity is 1. When the semantics of two word vectors are completely different, word similarity is close to 0. In this paper, the BioDesign model takes word similarity as an important indicator to evaluate whether word semantics are similar

3.3 Result analysis

The above content introduces the method of artificial intelligence to assist the collection of biomimetic concept information. However, in the process of biomimetic design, there are a lot of information to choose. For example, how to respond to the information summary and data involved in the biomimetic design in a short time is still a challenging task. Though natural language approaches to biomimetics usually search biological texts for words associated with engineeringthe results may not be the only descriptors of a biomimetic design. Considering the output results of the BioDesign model comprehensively, two variables are introduced in the result analysis: key information complexity K and key information elasticity weight a. The two variables are unutilized to analyze the key characteristics of the results and balance the interaction between multiple results. The formula for result analysis of BioDesign is as follows:

$$M(K_n, keyword_n) = x_{result_n} \tag{7}$$

The keyword is a key vocabulary that can express the concept of the subject of a document. Keyword extraction is the basic work of text retrieval, document difference comparison, abstract generation, text classification and clustering. M is the output of the corresponding BioDesign model. x_{result_n} is the response result of BioDesign model. Further, there has:

$$M(K_n, keyword_n) = \begin{bmatrix} M\left(K_1, keyword_1\right) \\ M\left(K_2, keyword_2\right) \\ \ldots \\ M\left(K_n, keyword_n\right) \end{bmatrix} = \begin{bmatrix} m_{11}, m_{12}, \ldots, m_{1Kn} \\ m_{21}, m_{22}, \ldots, m_{2Kn} \\ \ldots \\ m_{21}, m_{22}, \ldots, m_{2Kn} \end{bmatrix} \tag{8}$$

where, m is the element of the output result of the BioDesign model corresponding to K and keyword. In order to take into account the different importance of key information, the key information elasticity weight a is introduced:

$$a \cdot x_{result_n}^{T} = y \tag{9}$$

where, a is the elastic weight of key information. y is the corresponding elastic weight matrix and the biomimetic design result inspired by multiple sources. Further, there are:

$$a = a_{0n} + {a_{0n}}^{\theta} = [a_1, a_2, \ldots, a_n] \tag{10}$$

$$\theta = \frac{len[M(K_1, keyword_1)]}{len[M(K_1/2, keyword_1)]} \tag{11}$$

where, a_{0n} is the initial weight of the keyword. θ is the weight activity. $len[M(K_1/2, keyword_1)]$ is the number of elements in the model response result. The implicit function expression of the key information elasticity weight a on the variables and is shown in Figure 4. It is not difficult to see that after introducing weight activity variables, the range of key information weight changes has been significantly expanded. Therefore, the introduction of the key information elastic weight can widely consider the influence of the model output on the results of the multiple information source inspired by biomimetic design.

The pseudo code of the biomimetic inspired result analysis algorithm based on the BioDesign model is illustrated in Algorithm 1.

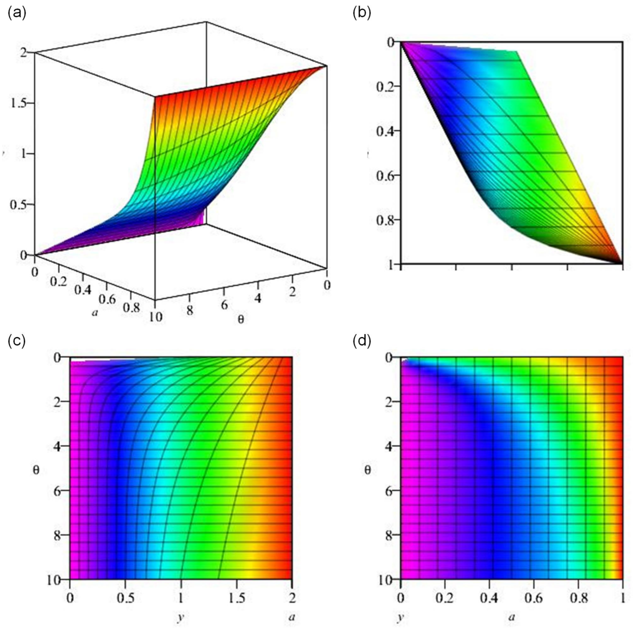

Fig. 4 The implicit function expression of the key information elasticity weight a with respect to the variables a_{0n} and θ.

Algorithm 1 Correlation analysis of biomimetic inspiration results from multiple sources

Input: Design the target keyword, the complexity of key information K, and the elastic matrix a
Output: y, Result of biomimetic design inspired by multiple sources

input K_n,$Keyword_n$, calculate x_{result_n}
input K_n, calculate $len[M(K_1, keyword_1)]$
input $K_n/2$, calculate $len[M(K_1/2, keyword_1)]$
return $\theta = \frac{len[M(K_1, keyword_1)]}{len[M(K_1/2, keyword_1)]}$
return $a = a_{0n} + {a_{0n}}^{\theta} = [a_1, a_2, \ldots, a_n]$
return quantitative recommended biological strategies

In the following part, the key information complexity K and key information elastic weight a are used to illustrate the impact of the above two variables on the results of the proposed BioDesign model in the biomimetic design system with the subject of erosion and wear related content. Using Python language programming to achieve the above algorithm. We used jupyter notebook as the development and debugging tools, running under the Windows 10 operating system.

(1) The influence of key information complexity K on the results.

The significance of correlation analysis: filter information and assist in processing information quickly. Therefore, the correlation analysis of the results is helpful to increase the information diversity of the biomimetic design and the ability of multi-dimensional rapid data processing. Figure 5 shows the output results corresponding to the input of the three key information

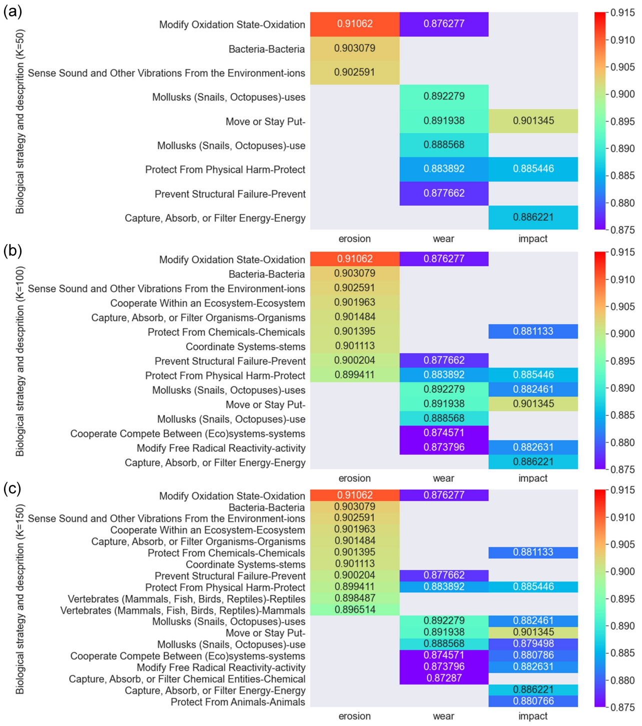

Fig. 5 BioDesign output results corresponding to key information "erosion", "wear" and "impact" with variable K: (a) $K = 50$, (b) $K = 100$, (c) $K = 150$.

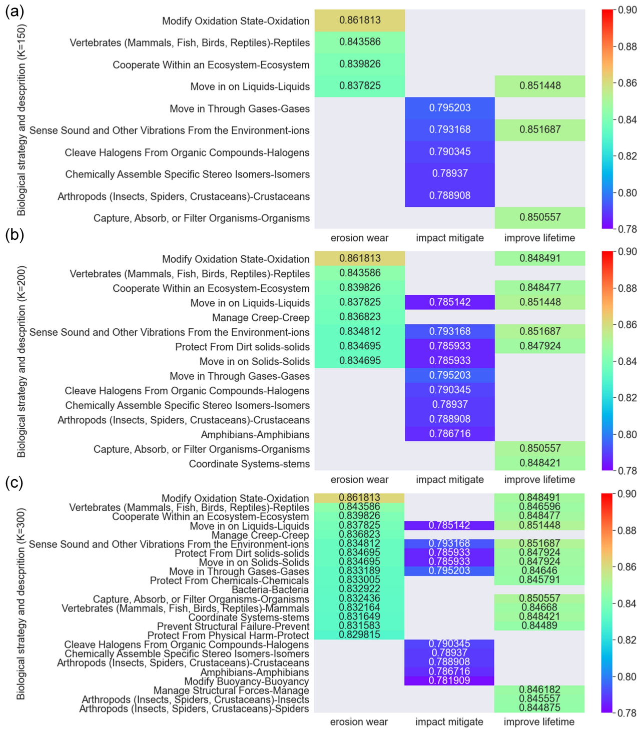

Fig. 6 BioDesign output results corresponding to key information "erosion wear", "impact mitigate", and "improve lifetime" with variable K: (a) $K = 150$, (b) $K = 200$, (c) $K = 300$.

"erosion", "wear" and "impact". We compared the output results of BioDesign at different K values and the same a. The heat map on the left can intuitively show the similarity between each input keyword and the returned result. As the K value increases, the complexity of the returned results gradually increases. At the same time, the commonality of the corresponding output results for each keyword also gradually increases. The right half of Figure 5 shows the Y value returned under the default elastic matrix and the corresponding biological strategy name and function description. It can be seen that the outputs of BioDesign increase with the increase of K. Similarly, we compared the output results corresponding to the key information of "erosion wear", "impact mitigate", and "improve lifetime". The results are shown in Figure 6. The law presented is the same as that shown in Figure 5 . It can be seen from the results in Figures 5 and 6 that the introduced K has a positive effect on the evaluation of BioDesign results. Thus it can be seen that K increase the range of retrieved biological functions.

(2) The relationship between key information elasticity weight a and results

In the previous part, we discussed BioDesign's help to biomimetic innovation and obtaining biological strategies based on key information. However, searching and understanding the results are still a time-consuming task. The introduction of a can help designers reduce the time to analyze results. In a general way, the weight of key information also has an important influence on the calculation results of the model. For example, when a designer wants to analyze multiple key information at the same time, due to different application environments, the weight of the key information may have different definitions. In order to analyze the impact of a_{0n} and θ on the results, we made the following analysis. Take the three keywords of "erosion", "wear" and "impact" as examples. Three different are initialized, as shown in Table 3. According to equation (11), the variable is related to the result of θ and K. Here we set $K = 200$, calculate $len[M(K_1, keyword_1)]$ and $len[M(K_1/2, keyword_1)]$ respectively, and find the corresponding 11/9, 9/7 and 9/3. Input the above parameters into the model and obtain the result output. In order to see the effect of a on the results more intuitively, the results are visualized, as shown in Figure 7. The vertical axis is the name and description of the analysis result. The horizontal axis is its corresponding similarity weighting result. The result of a is not considered as a control group. It can be seen from the results that, compared with the control group, the elasticity weight a can effectively highlight the results with greater sim-

ilarity. At the same time, it fully considers the role of key information of different complexity in the results.

Similarly, we used the same method to analyze the elastic matrix a of the three keyword groups "erosion wear", "impact mitigate", and "Improve lifetime" and the corresponding results as shown in Figure 8. As can be seen from the Figure 8 that a can respond to information about keywords whether it's a word or a phrase.

Table 3 The variable elastic weight (a) of key information: "erosion", "wear" and "impact".

a	keyword	a_{0n}	θ
	erosion	1	11/9
$a1$	wear	0.8	9/7
	impact	0.5	9/3
	erosion	1	11/9
$a2$	wear	0.1	9/7
	impact	0.5	9/3
	erosion	1	11/9
$a3$	wear	0.8	9/7
	impact	0.7	9/3

Table 4 The variable elastic weight (a) of key information: "erosion wear", "impact mitigate", and "improve lifetime".

a	keyword	a_{0n}	θ
	erosion wear	1	16/4
$a1$	impact mitigate	0.8	10/5
	improve lifetime	0.5	16/3
	erosion wear	1	16/4
$a2$	impact mitigate	0.1	10/5
	improve lifetime	0.5	16/3
	erosion wear	1	16/4
$a3$	impact mitigate	0.8	10/5
	improve lifetime	0.7	16/3

4 Case Study

The biologically inspired design method presented in this article is able to inspire more innovation. This case applies BioDesign model to biomimetic innovative design of erosion wear resistance of water hydraulic control valve.

4.1 Biomimetic innovation control valve core

One of the reasons behind failed engineering surfaces and mechanical components is particle erosion wear [37]. Erosion wear is a complex phenomenon [38], [39], [40], [41], [42], [43], [44], [45], [46] depends on erodent particle (size, shape, hardness, concentration), eroded substance (elastic properties, surface hardness, surface morphology), flow condition (impacting velocity, angle, location) and so on.

Therefore, the ideal structure that is resistant to erosion and wear should be able to resist erosion and wear, have a longer service life, and slow down particle impact and other properties. Therefore, in the erosion resistant design of the water hydraulic valve core, we hope that the valve core can be obtained the above advantages. Here, we choose three groups of keywords ("erosion wear", "impact mitigate", and "improve lifetime") as the input of BioDesign model.

A systematic research plan was developed to study the erosion wear of the control valve with biomimetic valve core. The following are the details of testing facilities and procedures. Refer to the application process of BioDesign mentioned above, first, determine the theme of designing the biomimetic anti-erosion valve core. Set the complexity $K = 300$. Combined with actual working conditions, in this case study, the initial weights of the three key information are given to the above keywords, which are 1, 0.1, and 0.3, respectively. Enter keywords into the model, and return $len[M(K_1, keyword_1)]$ and $len[M(K_1/2, keyword_1)]$ corresponding to the three key information. The calculation and model output results are shown in Figure 9. According to the top-down artificial intelligence-assisted biomimetic design method, we selected three biological strategies of "absorb", "move in solid", and "protect from dirt solid" from the results of biological strategies inspiration as multi-source biomimetic design inspiration. Then combined with the actual situation of the valve core, abstracted from the above inspiration as "absorb particle impact" and "move particle direction". Then we couple them with the spool to design a biomimetic erosion-resistant spool, as shown in Figure 9. We have improved the valve core structure and replaced part of the structure with a soft material (rubber layer) to achieve the effect of absorbing particle impact. At the same time, we set up non-smooth pits on the partial surface of the valve core. The fluid passes through the non-smooth surface and rotates in the pit, changing the flow field on the wall. The velocity of the vortex in the pit is opposite to the impact direction [47]. Achieved the purpose of "move particle direction".

4.2 Erosion wear experiment

In order to further analyze the validity of the biomimetic design results, we carried out erosion wear experiments on the valve core. The initial concentration (5 wt.% by weight) of sand particles were the same in all the erosion wear tests. The erosion wear experimental

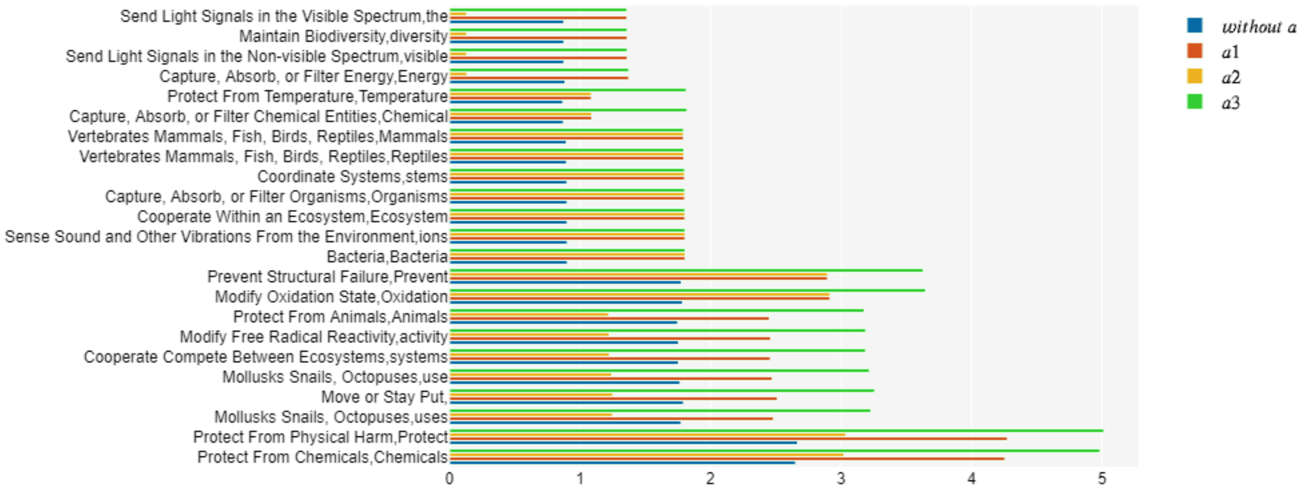

Fig. 7 BioDesign output results with variable elastic weight (a) of key information: "erosion", "wear", and "impact" at $K = 200$.

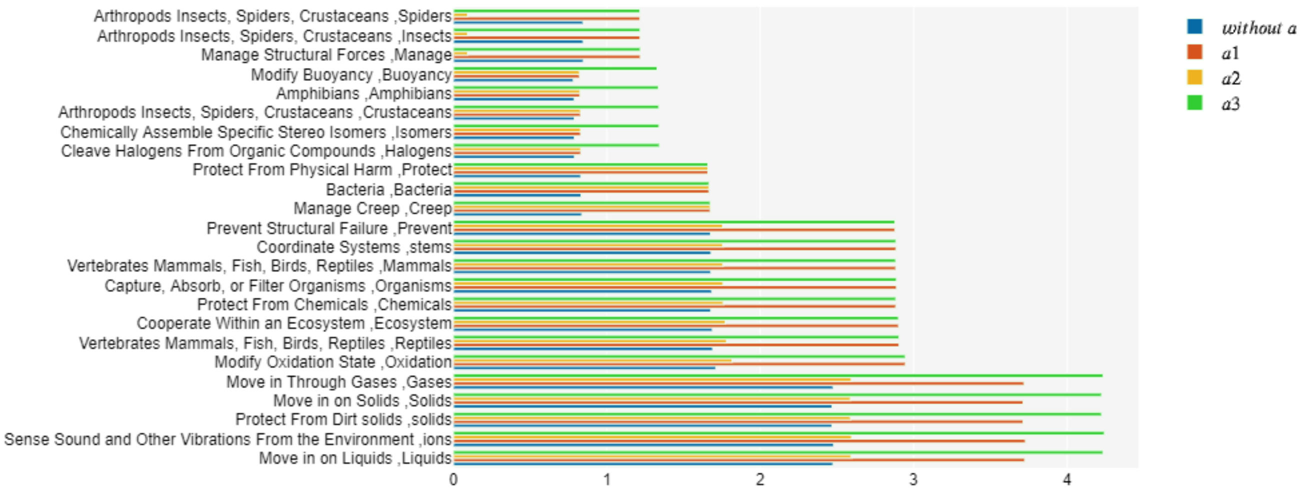

Fig. 8 BioDesign output results with variable elastic weight (a) of key information: "erosion wear", "impact mitigate", and "improve lifetime" at $K = 300$.

setup is schematically shown in Figure 10. Each group was tested at the same valve opening. Besides, the same flow is obtained by a PID closed-loop control algorithm. The control valve is calibrated to ensure the same initial position of valve displacement before each experiment. The control valve's opening is maintained at 8 mm by closed-loop control. All experiments in this study have been done on the 1 inch inside diameter (25.4 mm) pipe set. The experimental scheme of erosion wear is given in Table 5. The test valve core were fabricated by 3D printed, the material parameters of 3D printed valve plugs are illustrated in Table6. Among the erosion wear test, two factors were selected: A and B. Factor A examined two levels: smooth (1) and biomimetic geometry (2). Factor B examined two levels: no biomimetic soft layer (1) and biomimetic soft layer (2). Sand particles could remove the paint from highly erosion regions, therefore, the tested samples were sprayed black to observe the erosion paint removal form. All samples were put into the ultrasonic cleaning machine (Shenzhen Jiemong, 40 Hz) for 5 min after the erosion experiment. Then, dry the samples at room temperature for 12 hours. All samples were weighed with an electronic weighing balance (Sartorius, BSA224S, accuracy of 0.1 mg).

Table 5 Erosion wear experiment conditions.

parameter	value
test time	10 min
flow	13.5 L/min
particle diameter	0.3 mm
particle concentration	5 wt.%

For qualitatively compare the erosion resistance of the samples, the erosion rate is used as an evaluation standard. The erosion rate (ϵ) can be calculated as the following equation:

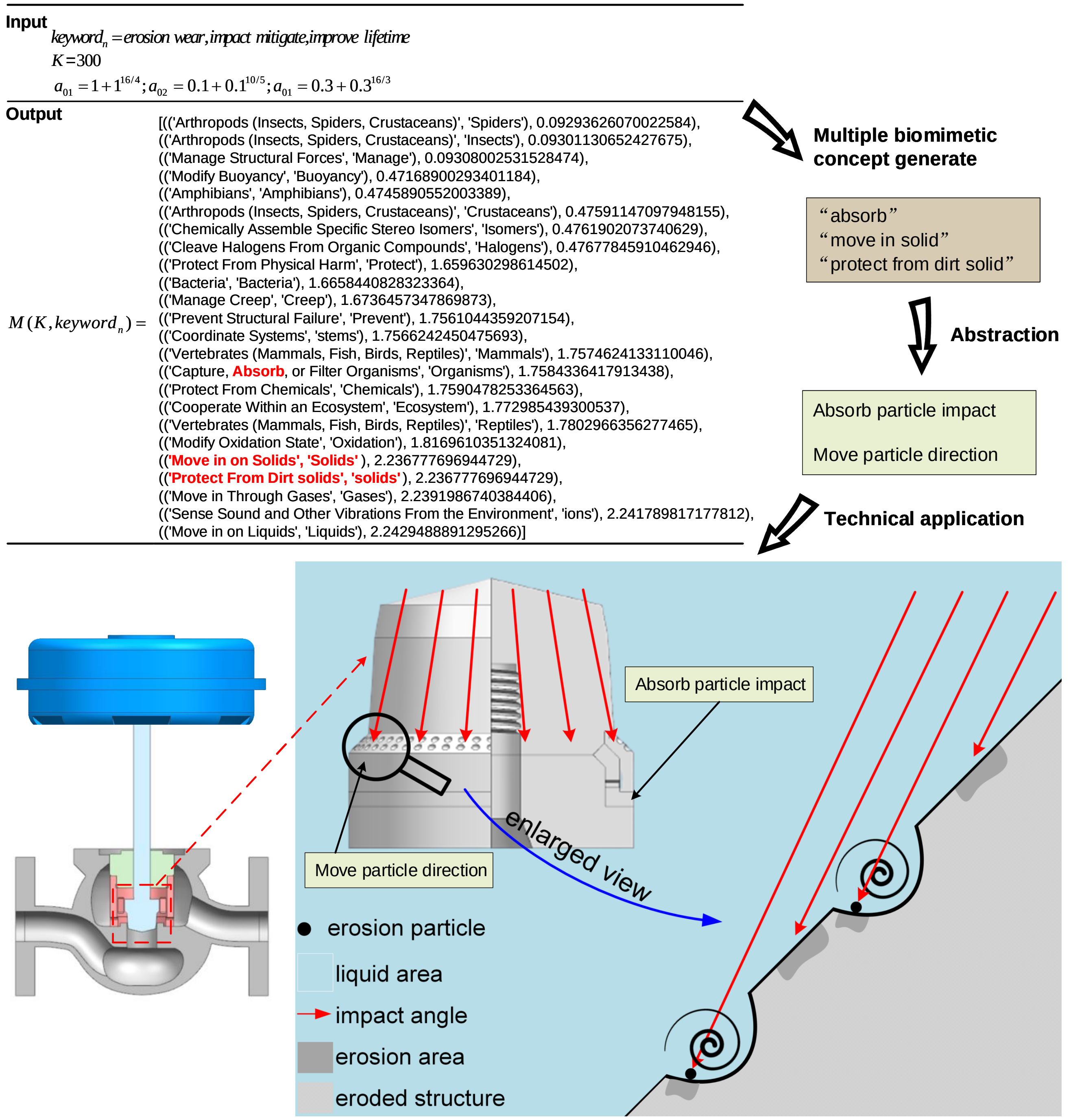

Fig. 9 The design out of erosion wear control valve core based on BioDesign.

Table 6 Material parameters of 3D printed valve plugs.

material	elastic modulus (MPa)	poisson's ratio	density (t/mm^3)
resin	2.5E+3	0.41	1.19E-9
rubber	-	0.49	1.2E-9

where, m_0 and m_1 are the mass of samples before and after experiment, t is the experiment period.

$$\epsilon = \frac{m_0 - m_1}{t} \tag{12}$$

For factor A, $\sum A_1 B_{all} = 7.6 \times 10^{-4}\ g/s$, $\sum A_2 B_{all} = 4.7 \times 10^{-4}\ g/s$. For factor B, $\sum A_{all} B_1 = 7.3 \times 10^{-4}\ g/s$, $\sum A_{all} B_2 = 5.0 \times 10^{-4}\ g/s$.

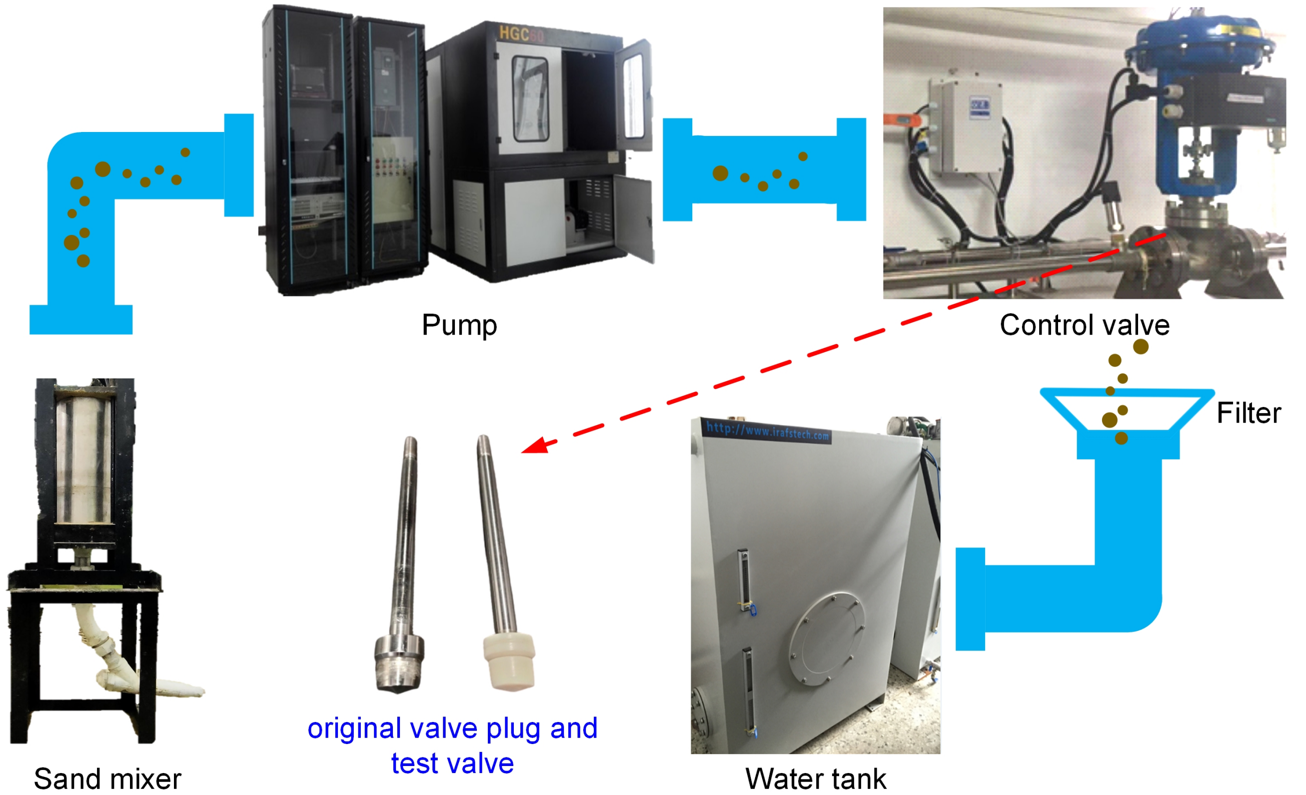

Fig. 10 Schematic and photograph of the experimental setup.

In other words, the valve core with biomimetic geometry has a lower erosion rate. The valve core with a soft layer show better erosion wear resistance, and the valve core has a soft layer with negative stiffness cell inside has the lowest erosion rate. Table 7 compares the erosion rate and paint removal form after 5 minutes of erosion wear. According to equation 12, Case 4 has a minimum erosion rate. Meanwhile, according to the erosion appearance, the erosion form of case 4 is also relatively low. That is the valve core with biomimetic geometry and soft layer shows the best erosion wear resistance performance.

In addition, in terms of erosion rate, the effect of non-smooth structure on erosion rate reduction is stronger than that of soft layer structure, which is consistent with the output similarity result of BioDesign model. In summary, according to the case analysis of erosion wear of valve core, the BioDesign model has a obvious positive effect on biomimetic design

Table 7 Analysis of the erosion wear test during 5 minutes.

case	A	B	erosion rate (10^{-4} g/s)	paint removal
1	1	1	4.4	
2	1	2	3.2	
3	2	1	2.9	
4	2	2	1.8	

5 Discussion and Conclusion

In the past few years, interdisciplinary research and development combining natural sciences (biology) and technology (design and engineering) have increased significantly. However, the limited personal perception, knowledge and experience ensure that the rationality of the biomimetic design is worthy of further discussion. In response to this problem, we tried to use artificial intelligence methods to give suggestions on biomimetic design and provide objective references for biomimetic design. This is different from directly searching for design-

related keywords, and also includes possible design information, reducing the subjectivity of biomimetic design and biological search time. However, deep learning models are also difficult to interpret, even for experienced practitioners. Even for deep models, the combination of expert knowledge and feature engineering can improve performance more than complex models. Therefore, the subject of this research is not to use this deep learning method to replace the classic method of biomimetic design but as an auxiliary means that can reduce the limitation of biomimetic design on the research direction of designers. Finally, a case study is used to illustrate the effectiveness of the BioDesign model in biomimetic design. By constructing the data-driven biomimetic design innovation model BioDesign, the driving force of biomimetic design innovation is improved, and design thinking innovation is promoted. The innovations of this article are:

(1) Based on the Asknature database, the content of biological strategy is used as an important reference for biomimetic design, and the complicated information is extracted into biological information and functional description.

(2) Innovative application of deep learning and natural language processing technology in biomimetic design inspiration recommendation. When using, you only need to enter the keywords of the design content you are interested in and the relevant keyword weights to obtain the recommended results and save time.

(3) The introduction of key information complexity indicators and the correlation analysis of the results expand the multi-dimensional direction of biomimetic design.

In this paper, we proposed a biological knowledge enabled BERT for innovation in biomimetic design. The proposed BioDesign can be utilized to improve the findability, recognizability and understandability in the process of biomimetic design. The BioDesign indicates the potential of biological knowledge-enabled NLP systems for systematic search strategies for biological inspiration:

Firstly, it is derived from the research data published by users for this biological-related function, and the collected perceptual words are more targeted;

Secondly, the data-oriented bionic evaluation method ensures the authenticity of the data;

Finally, collect data. The data collection efficiency is high, the large data volume and the real-time performance is good.

Of course, the proposed method has some limitations. In particular, building a deep learning model requires collecting a large amount of biomimetic design data. When the amount of data is insufficient, there are not many bionic design reference results returned. However, with the gradual development of the field of bionic design, this method can improve the efficiency of bionic conceptual design. In the future, exploring the cooperation between bionic experts and artificial intelligence may promote the development of the biomimetic design. When there is no bionic design expert, the designer can use this method to evaluate the engineering performance of the new design. Further development and broader evaluation of this method are part of future work.

6 Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgements This work was supported by the Natural Science Foundation of China under Grant 51875113, Natural Science Joint Guidance Foundation of the Heilongjiang Province of China under Grant LH2019E027. PhD Student Research and Innovation Fund of the Fundamental Research Funds for the Central Universities under Grant XK2070021009.

References

1. E. Cruz, T. Hubert, G. Chancoco, O. Naim, N. Chayaamor-Heil, R. Cornette, C. Menezo, L. Badarnah, K. Raskin, F. Aujard, Energy and Buildings **246**, 111034 (2021)
2. D. Palin, S. Russell, F.F.E. Kohle, E. O'Dowd, S.Y.T. Flynn, Dearq-Revista De Arquitectura-Journal of Architecture (26), 52 (2020)
3. T.N. Sullivan, T.T. Hung, A. Velasco-Hogan, M.A. Meyers, Materials Science & Engineering C-Materials for Biological Applications **105** (2019)
4. X.c. Zhang, C.c. An, Z.f. Shen, H.x. Wu, W.g. Yang, J.p. Bai, Materials Today Communications **23** (2020)
5. S. Bonfanti, R. Guerra, M. Zaiser, S. Zapperi, Apl Materials **9**(2) (2021)
6. O. Speck, D. Speck, R. Horn, J. Gantner, K.P. Sedlbauer, Bioinspiration & Biomimetics **12**(1) (2017)
7. R. Kruiper, J.F.V. Vincent, E. Abraham, R.C. Soar, I. Konstas, J. Chen-Burger, M.P.Y. Desmulliez, Biomimetics **3**(3) (2018)
8. E. Graeff, N. Maranzana, A. Aoussat, Journal of Engineering Design **30**(8-9), 289 (2019)
9. A. Chakrabarti, L. Siddharth, M. Dinakar, M. Panda, N. Palegar, S. Keshwani, *Idea Inspire 3.0-A Tool for Analogical Design* (2017), *Smart Innovation Systems and Technologies*, vol. 66, pp. 475–485
10. T.A. Lenau, A.L. Metze, T. Hesselberg, *Paradigms for biologically inspired design* (2018), *Proceedings of SPIE*, vol. 10593
11. D. Vandevenne, P.A. Verhaegen, S. Dewulf, J. Duflou, Artificial Intelligence for Engineering Design, Analysis and Manufacturing pp. 1–18 (2015)

12. H.H. Farzaneh, K. Helms, U. Lindemann, *VISUAL REPRESENTATIONS AS A BRIDGE FOR ENGINEERS AND BIOLOGISTS IN BIO-INSPIRED DESIGN COLLABORATIONS* (2015). International Conference on Engineering Design
13. X. Wang, X. Yang, J. Du, X. Wang, J. Li, X. Tang, Scientometrics **126**(7), 5531 (2021)
14. D. Ofer, N. Brandes, M. Linial, Computational and Structural Biotechnology Journal **19**, 1750 (2021)
15. X. Li, B. Lin, Mathematical Problems in Engineering **2021** (2021)
16. W. Sha, Y. Guo, Q. Yuan, S. Tang, X. Zhang, S. Lu, X. Guo, Y.C. Cao, S. Cheng, Advanced Intelligent Systems **2**(4) (2020)
17. R. Verganti, L. Vendraminelli, M. Iansiti, Journal of Product Innovation Management **37**(3), 212 (2020)
18. F.S. Zhong, J. Xing, X.T. Li, X.H. Liu, Z.Y. Fu, Z.P. Xiong, D. Lu, X.L. Wu, J.H. Zhao, X.Q. Tan, F. Li, X.M. Luo, Z.J. Li, K.X. Chen, M.Y. Zheng, H.L. Jiang, Science China-Life Sciences **61**(10), 1191 (2018)
19. G.X. Gu, C.T. Chen, D.J. Richmond, M.J. Buehler, Materials Horizons **5**(5), 939 (2018)
20. J.L. Krichmar, W. Severa, M.S. Khan, J.L. Olds, Frontiers in Neuroscience **13** (2019)
21. A. Buche, M.B. Chandak, Bioscience Biotechnology Research Communications **13**(14), 35 (2020)
22. S. Nam, S. Yoon, N. Raghavan, H. Park, Sustainability **13**(1) (2021)
23. T. Chen, M. Wu, H. Li, Database-the Journal of Biological Databases and Curation (2019)
24. J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C.H. So, J. Kang, Bioinformatics **36**(4), 1234 (2020)
25. J. Wang, X. Zhang, L. Chen, Knowledge-Based Systems **232** (2021)
26. O. Speck, D. Speck, R. Horn, J. Gantner, K.P. Sedlbauer, Bioinspir Biomim **12**(1), 011004 (2017)
27. E. Graeff, N. Maranzana, A. Aoussat, in *22th International Conference on Engineering Design (ICED19), Delft, The Netherlands*
28. M. Willocx, A. Ayali, J.R. Duflou, Cirp Journal of Manufacturing Science and Technology **31**, 61 (2020)
29. T. Lenau, A.L. Metze, T. Hesselberg, *Paradigms for biologically inspired design*, *SPIE Smart Structures and Materials + Nondestructive Evaluation and Health Monitoring*, vol. 10593 (SPIE, 2018)
30. K. Fu, D. Moreno, M. Yang, K.L. Wood, Journal of Mechanical Design **136**(11) (2014)
31. J. Sartori, U. Pal, A. Chakrabarti, Artificial Intelligence for Engineering Design, Analysis and Manufacturing **24**, 483 (2010)
32. C.T.A. Pham, S. Magistretti, C. Dell'Era, Innovation-Organization & Management p. 24 (2021)
33. N. Thuethongchai, T. Taiphapoon, A. Chandrachai, S. Triukose, Social Sciences-Basel **9**(3) (2020)
34. J. Devlin, M.W. Chang, K. Lee, K. Toutanova, arXiv preprint arXiv:1810.04805 (2018)
35. T. Saha, S.R. Jayashree, S. Saha, P. Bhattacharyya, Ieee Transactions on Computational Social Systems **7**(5), 1168 (2020)
36. M. Pota, M. Ventura, R. Catelli, M. Esposito, Sensors **21**(1) (2021)
37. F. Sun, H. Xu, Bio-Design and Manufacturing **3**(4), 331 (2020)
38. M. Ahmad, M. Schatz, M. Casey, Wear **303**(1-2), 83 (2013)
39. A.P. Anaraki, J. Kadkhodapour, B. Taherkhani, Journal of Failure Analysis and Prevention **14**(6), 784 (2014)
40. P. Balu, F. Kong, S. Hamid, R. Kovacevic, Tribology International **62**, 18 (2013)
41. S. Hassani, M. Bielawski, W. Beres, L. Martinu, M. Balazinski, J. Klemberg-Sapieha, Surface and Coatings Technology **203**(3-4), 204 (2008)
42. J. Kadkhodapour, A.P. Anaraki, B. Taherkhani, Journal of Failure Analysis and Prevention **15**(2), 272 (2015)
43. H.A. Sherif, F.A. Almufadi, Wear **412-413**, 127 (2018)
44. B.S. Thapa, B. Thapa, O.G. Dahlhaug, Energy **41**(1), 386 (2012)
45. B. Yıldırım, S. Müftü, International Journal of Solids and Structures **49**(11-12), 1375 (2012)
46. N. Zhang, F. Yang, L. Li, C. Shen, J. Castro, L.J. Lee, Wear **303**(1-2), 49 (2013)
47. Z.W. Han, J.Q. Zhang, C. Ge, L. Wen, L.Q. Ren, Langmuir **28**(5), 2914 (2012)